Test and Evaluation Policies and Practices: A New Emphasis

Darlene Mosser-Kerner

Office of the Deputy Director,
Developmental Test and Evaluation,
Office of the Secretary of Defense, Washington, D.C.

The Department of Defense (DoD) recently issued new and revised test and evaluation (T&E) policies that represent a shift in emphasis toward the evaluation side of T&E and promote a continued emphasis on integrated testing. The revised policies focus on using T&E throughout the system life cycle in a seamless continuum. This revision of T&E policies represents one of many actions the Department is taking to revitalize T&E and to ensure that the T&E is timely, effective, and efficient.

Key words: 231 report; capabilities and limitations; Department of Defense; evaluation and reporting; system life cycle; test and evaluation.

In December 2007, the Under Secretary of Defense for Acquisition, Technology, and Logistics, and the Director of Operational Test and Evaluation jointly issued a memo to introduce new and revised policies for test and evaluation (T&E) of Department of Defense (DoD) programs. The memo affirms, "The fundamental purpose of test and evaluation is to provide knowledge to assist in managing the risks involved in developing, producing, operating, and sustaining systems and capabilities" (OSD 2007).

The revised policy responds to a 2007 review of DoD T&E and its applicability to emerging acquisition approaches. The Director of Operational Test and Evaluation and the Office of the Deputy Director, Developmental Test and Evaluation conducted the review and delivered the resulting report to Congress in July 2007 in compliance with Section 231 of the John Warner National Defense Authorization Act for Fiscal Year 2007, Public Law 109-364. The report, known as the "231 report," is the latest in a series of reviews and studies of DoD T&E that signaled the shift in DoD T&E policy.

The December 2007 policy and the findings from the 231 report can be grouped into four broad themes:

1. Emphasis on evaluation
2. Focus on capabilities and limitations
3. Integrated and seamless T&E
4. Developmental T&E reporting.

Emphasis on evaluation

In recent years the Department has focused on the testing side of T&E, creating an imbalance toward measuring technical parameters, but the new policy assumes the "knowledge to assist in managing risk" (OSD 2007) comes mainly from the evaluation step of the T&E process. Testing is perhaps the most visible part of T&E and consumes most of the resources; however, people conduct testing because someone in a decision-making role needs credible knowledge of how a system works or does not work to make an informed decision.

The effectiveness of the evaluation depends on decisions about what to test and the applicability of the data from testing. If program managers assume that they cannot test all aspects of a system or capability, then the questions become twofold: What do they test, and how much testing is enough? The answer at a strategic level is to test enough, and in specific areas, to mitigate the key risks for the system or capability being developed.

Who defines the key risks? The program manager for one, and all the decision makers in the program management chain, which includes the milestone decision authorities, and even Congress, which authorizes and appropriates funding for the program. Other "decision makers" who need T&E-generated knowledge to manage risk include systems engineers who need knowledge of system and subsystem performance to assist in maturing the technologies and design. The

Report Documentation Page		*Form Approved* *OMB No. 0704-0188*

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **JUN 2008**	2. REPORT TYPE	3. DATES COVERED **00-00-2008 to 00-00-2008**
4. TITLE AND SUBTITLE **Test and Evaluation Policies and Practices: A New Emphasis**		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Office of the Secretary of Defense,Developmental Test and Evaluation,Washington,DC,20301**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**		
13. SUPPLEMENTARY NOTES		
14. ABSTRACT		
15. SUBJECT TERMS		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**3**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

manufacturing decision makers need knowledge of system performance to mature and control the manufacturing processes. The operator uses the knowledge of system capabilities and limitations to mitigate the inherent risks in operating and employing the equipment. The maintainers need knowledge to inspect, service, and repair systems.

How much testing is enough? The obvious answer is "it depends." It depends upon how much risk the decision maker is willing to accept. If the decision maker is not willing to accept much risk, then the amount of required testing will increase. If the decision maker is willing to accept more risk, then the amount of required testing will decrease. In general, the expectation is that you will never have enough time or money to test to achieve absolute certainty; there will always be an element of uncertainty or residual risk.

By shifting the emphasis to evaluation and the knowledge generated through T&E, the customers are empowered, the decision makers are empowered to help testers determine what to test and how much testing is necessary. In some respects, this shift in emphasis will increase the importance of communication between the T&E community and the various decision makers.

Focus on capabilities and limitations

The second theme of the new policy on T&E is the focus on determining or assessing capabilities and limitations of the system(s). One of the purposes of the Defense acquisition system is to "acquire quality products that satisfy user needs with measurable improvements to mission capability" (DoDI 5000.1). One of the new policies is that "Evaluations shall include a comparison with current mission capabilities…, so that measurable improvements can be determined" (OSD 2007). This policy statement was driven by the use of relative performance in system requirements and during milestone reviews. For example, "System X shall be twice as good as Legacy Y," or "I know it doesn't meet the users' requirements, but it's better than what they currently have." The new policy recognizes the use and utility of comparative assessments and provides some appropriate guidance for the acquisition community.

In addition, the policy revision states that these improvements to mission capability "should be reported in terms of operational significance to the user" (OSD 2007). The focus on determining capabilities and limitations is not a mandate or a blank check to test everything in a search for capability or potential limitations. The amount of testing is still bounded by the risk tolerance of the various decision makers, especially the ones paying for the program. On the other hand, the focus on capabilities and limitations also means that T&E is more than just specification compliance. T&E does measure progress in system and capability development, and one of the ways to do that is by measuring progress against the specification; however, T&E should also develop an understanding of basic capabilities and limitations, so the systems engineer and the program manager can both assess the relative technical maturity of the system. The understanding of capabilities and limitations informs discussions of current mission performance and potential issuance of new capability requirements. The results of T&E need to be linked in some mission context and stated in terms of relevance to the user. Our purpose in defense acquisition is to provide capability to the user, so it makes sense that evaluators should be able to tie the results of T&E to capability for the user.

The focus on capabilities and limitations generated considerable discussion during the drafting and coordination of both the 231 report and the policy memorandum. The concern was specifically about the requirement to compare the new system capabilities with current mission capabilities and whether that requirement became an "unfunded mandate" to retest legacy systems. Such a mandate was not the intent, and the policy memo specifically included a provision that if the "evaluation is considered cost prohibitive the Service Component shall propose an alternative evaluation strategy" (OSD 2007). The new policy let the program managers know that if they wanted to use the rationale that the new system was better than the old system, they would need to provide a basis for that evaluation.

Integrated and seamless T&E

The third theme of the new policy is integrated and seamless T&E, meaning T&E conducted in a continuum throughout the system life cycle. The traditional focus of T&E has been during the system development phase and early production. One focus of the new policy is getting the T&E community involved earlier in the system life cycle, when requirements and concepts are first developed. The goals of this early involvement are to establish better requirements that are more fully understood, and the "early identification of technical, operational, and system deficiencies, so that appropriate and timely corrective actions can be developed prior to fielding the system" (OSD 2007).

In addition, "Developmental and operational test activities shall be integrated and seamless throughout the system life cycle" (OSD 2007). The focus on integrated developmental and operational testing is consistent with prior policy; however, now the role of T&E in the system life cycle is being expanded, so all

testing should be as seamless as possible, with minimal or no stops and starts for different types of testing. This seamless T&E will require continued emphasis on the use of live, virtual, and constructive modeling and simulation (M&S), or as the policy memo puts it, "T&E will be conducted in a continuum of live, virtual, and constructive system and operational environments" (OSD 2007). Another focus in making T&E integrated and more efficient is the policy that "evaluations shall take into account all available and relevant data from contractor and government sources" (OSD 2007). This may not be as easy as it sounds, given the typical issues with data authentication, archival, and retrieval, in addition to potential proprietary issues; however, it is essential if programs are to realize the promise of integrated testing in increasing the efficiency of the test programs and effectively shortening the time required to acquire new or improved capabilities for the warfighter.

T&E also should consider the deployment and sustainment period in the system life cycle. The new policy states in part, "As technology, software, and threats change, follow-on T&E should be used to assess current mission performance and inform operational users' during the development of new capability requirements" (OSD 2007). Since the majority of the life of a system is spent in operations and sustainment, T&E will have a role to play in providing system modifications, and assessments for end-of-life and disposal decisions. Some of the testing in this phase of the system life cycle is already being performed by operational units, so the new policy should not change that testing; however, it should cause a reassessment of all T&E throughout the system life cycle to ensure the full benefits of T&E are being realized in an efficient and effective manner.

Developmental T&E reporting

The fourth theme of the T&E policy memorandum is the renewed emphasis on evaluation and reporting by the developmental evaluators. This is one of the key aspects in revitalizing T&E, especially the government's Developmental Test & Evaluation role and mission. The operational evaluators already fulfill their statutory roles in providing assessments of operational effectiveness and suitability. In a similar manner, the developmental evaluators formerly provided assessments of system maturity and technical progress at each milestone decision review, but over the years that assessment has been lost. The new policy provides for a developmental evaluation of system "strengths and weaknesses in meeting the warfighters' documented

needs" (OSD 2007). The program manager is tasked with providing the results of this evaluation at the Milestone B and C reviews, so the new policy just adds a new element to the program manager's presentation. It does not create any additional independent reporting requirement.

Summary

The 231 report and associated policy memorandum are not the last word in revitalizing T&E in DoD. The Department is taking ongoing actions, in areas such as system of systems T&E for example, to revitalize the role T&E plays in the acquisition of new and modified systems and capabilities. The revised policy does provide a shift in emphasis on the role of T&E, and especially evaluations. The 231 report and policy memo also make adjustments in T&E policy to accommodate both existing and emerging acquisition approaches. The revised policy is another step toward achieving the end goal of efficient and effective testing to deliver timely knowledge to all stakeholders to help manage the risks in developing, producing, operating, and sustaining systems and capabilities for the Department of Defense. ❏

DARLENE MOSSER-KERNER is the senior policy lead within the Office of the Secretary of Defense for Developmental Test & Evaluation matters. She has more than 20 years of test and evaluation experience at NASA and the Department of Defense, including chief engineer on the NASA Systems Research Aircraft, assistant senior technical advisor at Edwards Air Force Base, and assistant program manager for T&E with the Naval Air Systems Command. She is the recipient of an Aviation Week & Technology Laurels Award and was the 2007 Department of Defense Civilian Tester of the Year. She has a bachelor's degree in electrical engineering and a master's degree in technology management. E-mail: Darlene.Mosser–Kerner@osd.mil.

References

Department of Defense. Department of Defense Directive 5000.1: The Defense Acquisition System. U. S. Department of Defense, 12 May 2003.

Department of Defense. Department of Defense Report to Congress on Policies and Practices for Test and Evaluation, July 2007.

Office of the Secretary of Defense Memorandum. Test and Evaluation Policy Revisions, 22 December 2007.